

Object detection on aerial images

Malyutina Ekaterina
(GeoScan)

Task overview



Prepare data

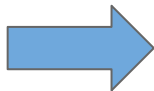
Full image
(4000 x 6000)

Train data
(640 x 480)

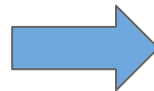


Classical pipeline

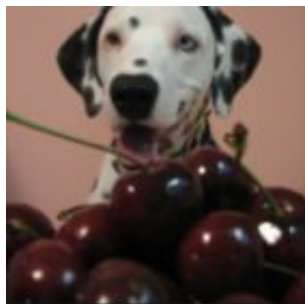
Classification



Localization

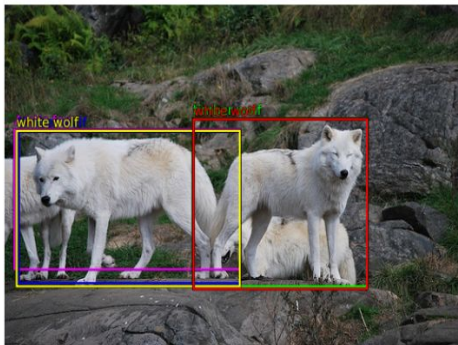


Detection



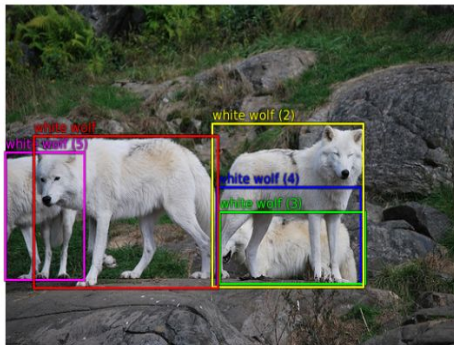
cherry

dalmatian
grape
elderberry
ffordshire bullterrier
currant



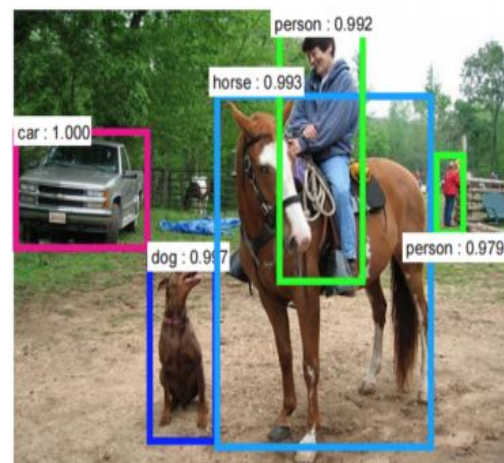
Top 5:

white wolf
white wolf
timber wolf
timber wolf
Arctic fox



Groundtruth:

white wolf
white wolf (2)
white wolf (3)
white wolf (4)
white wolf (5)



Our pipeline



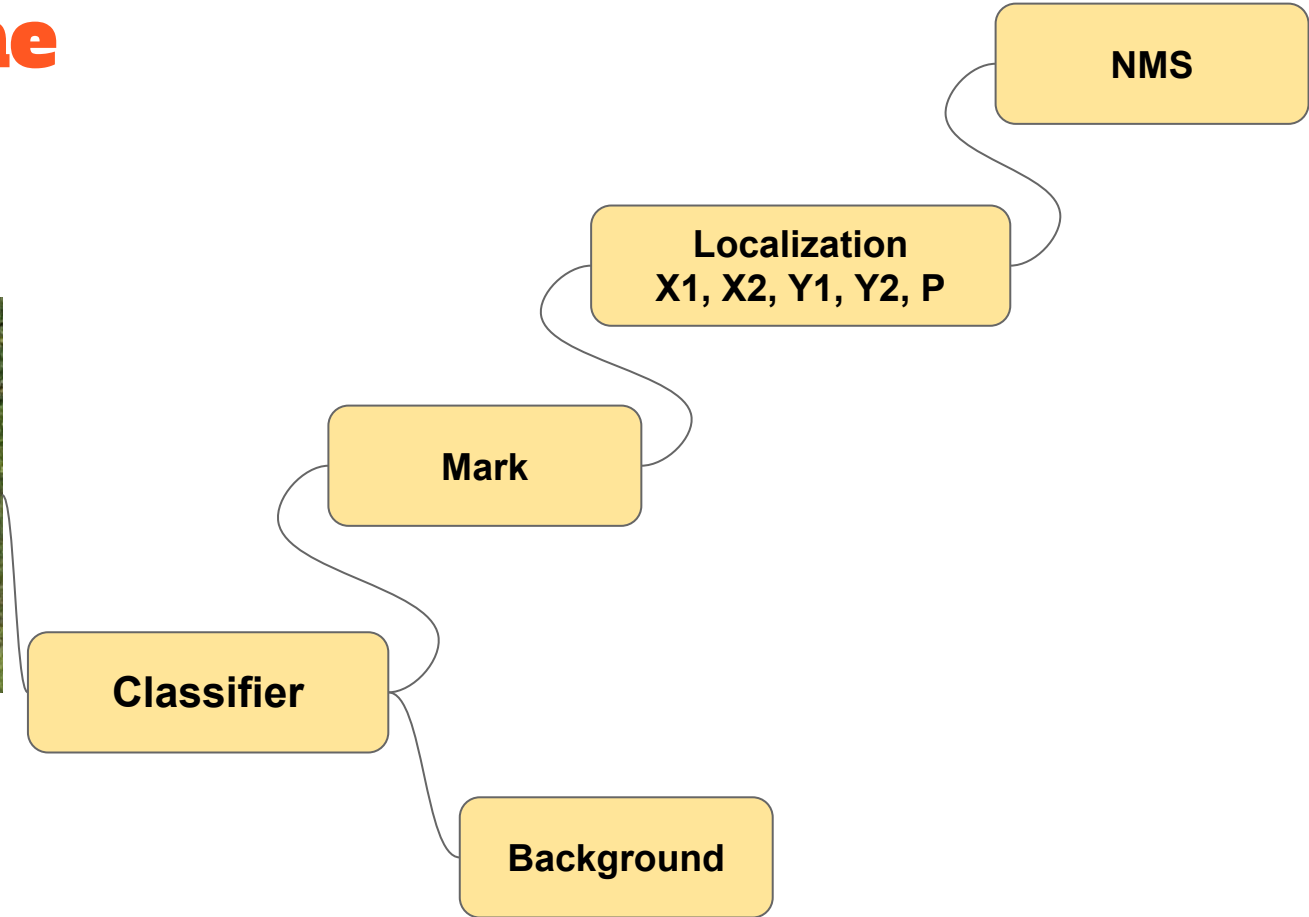
Classifier

Mark

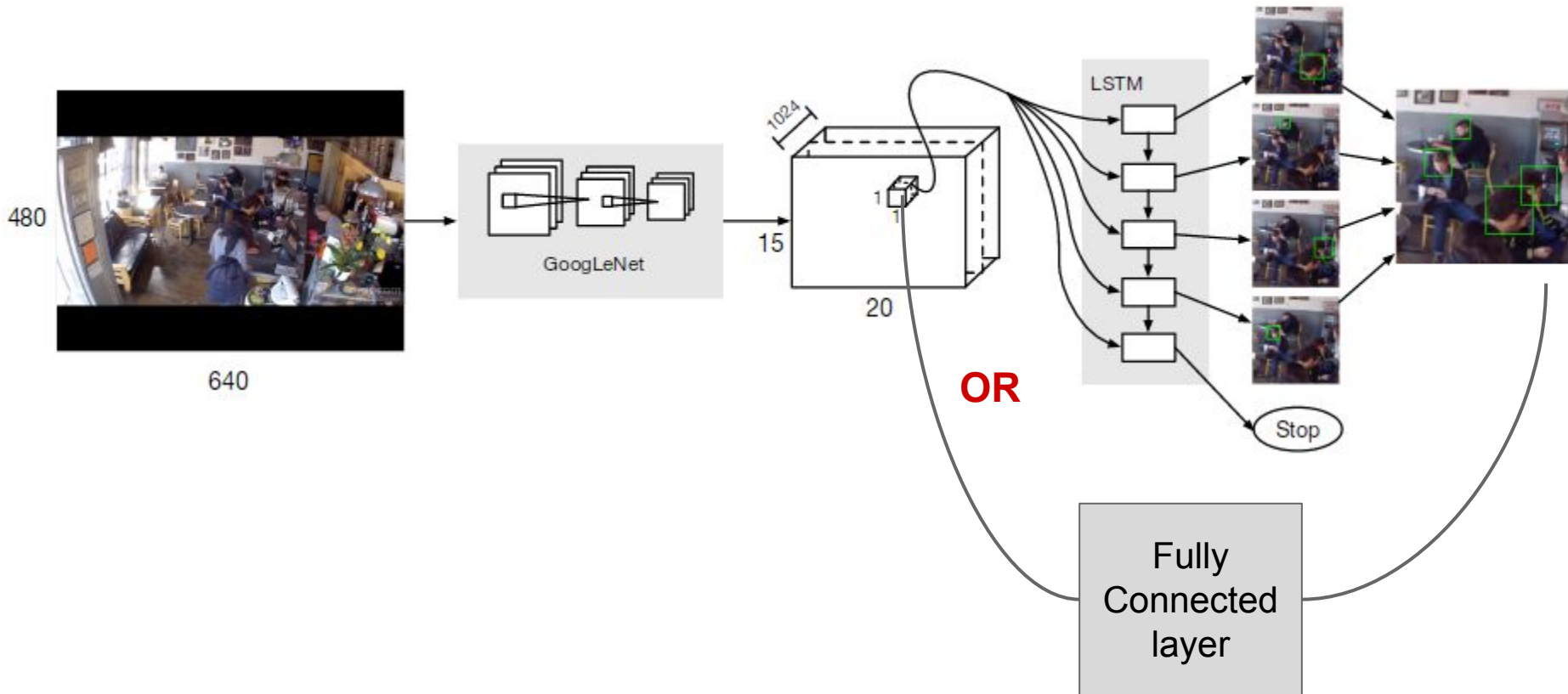
Localization
X1, X2, Y1, Y2, P

NMS

Background



More detailed lookup



Model training

- Weights initialized from GoogLeNet
- Loss function

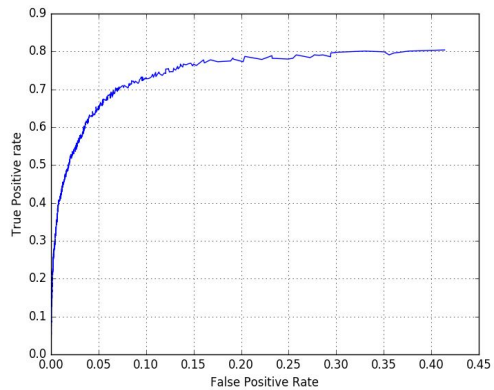
$$L(G, C) = \alpha \sum_{i=1}^{|G|} \left\| (b_i^{pos}, -\hat{b}_i^{pos}) \right\|_1 + \sum_{j=1}^{|C|} \text{CrossEntropy}(b_c^j, y_j)$$

$b = \{b_{pos} = \{b_x, b_y, b_w, b_h\}, b_c \in \{0, 1\}\}$ - Model output

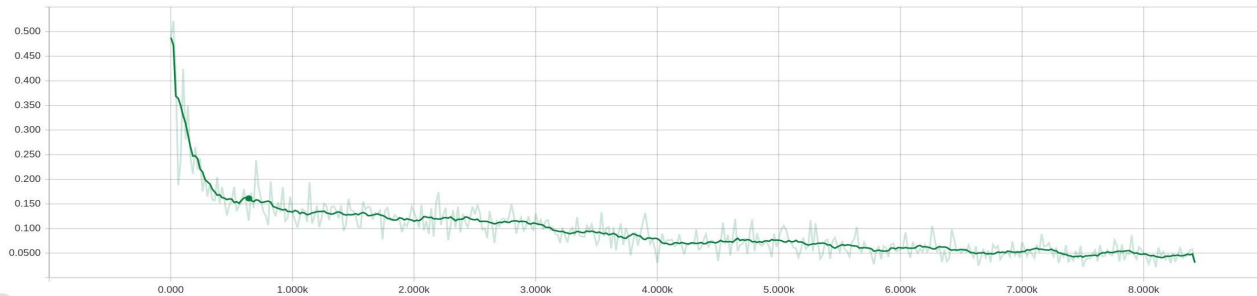
$G = \{b_i | i = 1 \dots L\}, C = \{\tilde{b}_i | i = 1 \dots N\}$ - Ground-truth and candidate bounding boxes respectively

- Optimizer: RMSPropOptimizer(learning rate = 0.001, epsilon = 1e-5)

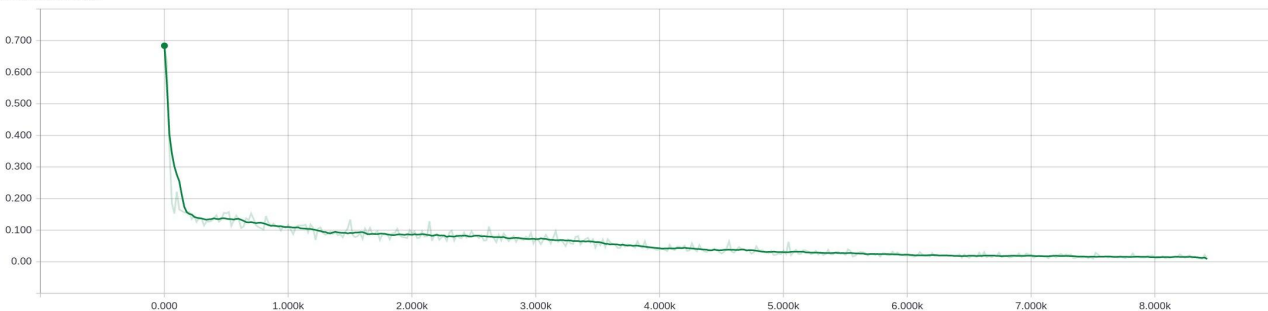
Results



test/regression_loss



test/confidences_loss



Results









References

- [End-to-end people detection in crowded scenes](#)
- [OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks](#)